Generative Self-training for Cross-domain Unsupervised Tagged-to-Cine MRI Synthesis

Xiaofeng Liu[1], **Fangxu Xing**[1], **Maureen Stone**[2], **Jiachen Zhuo**[2], **Timothy Reese**[3], **Jerry L. Prince**[4], **Georges El Fakhri**[1], **Jonghye Woo**[1]

[1]Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114, USA

[2]Dept. of Neural and Pain Sciences, University of Maryland School of Dentistry, Baltimore, MD, USA

[3]Athinoula A. Martinos Center for Biomedical Imaging, Dept. of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[4]Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Abstract

Self-training based unsupervised domain adaptation (UDA) has shown great potential to address the problem of domain shift, when applying a trained deep learning model in a source domain to unlabeled target domains. However, while the self-training UDA has demonstrated its effectiveness on discriminative tasks, such as classification and segmentation, via the reliable pseudo-label selection based on the softmax discrete histogram, the self-training UDA for generative tasks, such as image synthesis, is not fully investigated. In this work, we propose a novel generative self-training (GST) UDA framework with continuous value prediction and regression objective for cross-domain image synthesis. Specifically, we propose to filter the pseudo-label with an uncertainty mask, and quantify the predictive confidence of generated images with practical variational Bayes learning. The fast test-time adaptation is achieved by a round-based alternative optimization scheme. We validated our framework on the tagged-to-cine magnetic resonance imaging (MRI) synthesis problem, where datasets in the source and target domains were acquired from different scanners or centers. Extensive validations were carried out to verify our framework against popular adversarial training UDA methods. Results show that our GST, with tagged MRI of test subjects in new target domains, improved the synthesis quality by a large margin, compared with the adversarial training UDA methods.

1 Introduction

Deep learning has advanced state-of-the-art machine learning approaches and excelled at learning representations suitable for numerous discriminative and generative tasks [29, 22,

---

[5]It can be rewritten as $min_{\mathbf{w}}\mathscr{F} = \left\{ \sum_{t \in T} \sum_{n=1}^{N} \frac{1}{\sigma_{t,n}^2} \|(\hat{y}_{t,n} - \tilde{y}_{t,n})m_{t,n}\|_2^2 + \beta\left(\sum_{t \in T}\sum_{n=1}^{N} log\sigma_{t,n}^2 - C\right)\right\}$. Since β, C 0, an upper bound on $\mathscr{F}$ can be obtained as $\mathscr{F} \leq \mathscr{L}_{reg}^t$.

14, 21]. However, a deep learning model trained on labeled data from a source domain, in general, performs poorly on unlabeled data from unseen target domains, partly because of discrepancies between source and target data distributions, i.e., domain shift [15]. The problem of domain shift in medical imaging arises, because data are often acquired from different scanners, protocols, or centers [17]. This issue has motivated many researchers to investigate unsupervised domain adaptation (UDA), which aims to transfer knowledge learned from a labeled source domain to different but related unlabeled target domains [30, 33].

There has been a great deal of work to alleviate the domain shift using UDA [30]. Early methods attempted to learn domain-invariant representations or to take instance importance into consideration to bridge the gap between the source and target domains. In addition, due to the ability of deep learning to disentangle explanatory factors of variations, efforts have been made to learn more transferable features. Recent works in UDA incorporated discrepancy measures into network architectures to align feature distributions between source and target domains [18, 19]. This was achieved by either minimizing the distribution discrepancy between feature distribution statistics, e.g., maximum mean discrepancy (MMD), or adversarially learning the feature representations to fool a domain classifier in a two-player minimax game [18].

Recently, self-training based UDA presents a powerful means to counter unknown labels in the target domain [33], surpassing the adversarial learning-based methods in many discriminative UDA benchmarks, e.g., classification and segmentation (i.e., pixel-wise classification) [31, 23, 26]. The core idea behind the deep self-training based UDA is to iteratively generate a set of one-hot (or smoothed) pseudo-labels in the target domain, followed by retraining the network based on these pseudo-labels with target data [33]. Since outputs of the previous round can be noisy, it is critical to only select the high confidence prediction as reliable pseudo-label. In discriminative self-training with softmax output unit and cross-entropy objective, it is natural to define the confidence for a sample as the max of its output softmax probabilities [33]. Calibrating the uncertainty of the regression task, however, can be more challenging. Because of the insufficient target data and unreliable pseudo-labels, there can be both *epistemic* and *aleatoric* uncertainties [3] in self-training UDA. In addition, while the self-training UDA has demonstrated its effectiveness on classification and segmentation, via the reliable pseudo-label selection based on the softmax discrete histogram, the same approach for generative tasks, such as image synthesis, is underexplored.

In this work, we propose a novel generative self-training (GST) UDA framework with continuous value prediction and regression objective for tagged-to-cine magnetic resonance (MR) image synthesis. More specifically, we propose to filter the pseudo-label with an uncertainty mask, and quantify the predictive confidence of generated images with practical variational Bayes learning. The fast test-time adaptation is achieved by a round-based alternative optimization scheme. Our contributions are summarized as follows:

- We propose to achieve cross-scanner and cross-center test-time UDA of tagged-to-cine MR image synthesis, which can potentially reduce the extra cine MRI acquisition time and cost.

- A novel GST UDA scheme is proposed, which controls the confident pseudo-label (continuous value) selection with a practical Bayesian uncertainty mask. Both the aleatoric and epistemic uncertainties in GST UDA are investigated.

- Both quantitative and qualitative evaluation results, using a total of 1,768 paired slices of tagged and cine MRI from the source domain and tagged MR slices of target subjects from the cross-scanner and cross-center target domain, demonstrate the validity of our proposed GST framework and its superiority to conventional adversarial training based UDA methods.

2 Methodology

In our setting of the UDA image synthesis, we have paired resized tagged MR images, $\mathbf{x}_s \in \mathbb{R}^{256 \times 256}$, and cine MR images, $\mathbf{y}_s \in \mathbb{R}^{256 \times 256}$, indexed by $s = 1, 2, \cdots, S$, from the source domain $\{\mathbf{X}_S, \mathbf{Y}_S\}$, and target samples $\mathbf{x}_t \in \mathbb{R}^{256 \times 256}$ from the unlabeled target domain $\mathbf{X}_T$, indexed by $t = 1, 2, \cdots, T$. In both training and testing, the ground-truth target labels, i.e., cine MR images in the target domain, are inaccessible, and the pseudo-label $\widehat{\mathbf{y}}_t \in \mathbb{R}^{256 \times 256}$ of $\mathbf{x}_t$ is iteratively generated in a self-training scheme [33, 16]. In this work, we adopt the U-Net-based Pix2Pix [9] as our translator backbone, and initialize the network parameters $\mathbf{w}$ with the pre-training using the labeled source domain $\{\mathbf{X}_S, \mathbf{Y}_S\}$. In what follows, alternative optimization based self-training is applied to gradually update the U-Net part for the target domain image synthesis by training on both $\{\mathbf{X}_S, \mathbf{Y}_S\}$ and $\mathbf{X}_T$. Fig. 1 illustrates the proposed algorithm flow, which is detailed below.

2.1 Generative Self-training UDA

The conventional self-training regards the pseudo-label $\widehat{\mathbf{y}}_t$ as a learnable latent variable in the form of a categorical histogram, and assigns all-zero vector label for the uncertain samples or pixels to filter them out for loss calculation [33, 16]. Since not all pseudo-labels are reliable, we define a confidence threshold to progressively select confident pseudo-labels [32]. This is akin to self-paced learning that learns samples in an easy-to-hard order [12, 27]. In classification or segmentation tasks, the confidence can be simply measured by the maximum softmax output histogram probability [33]. The output of a generation task, however, is continuous values and thus setting the pseudo-label as 0 cannot drop the uncertain sample in the regression loss calculation.

Therefore, we first propose to formulate the generative self-training as a unified regression loss minimization scheme, where pseudo-labels can be a pixel-wise continuous value and indicate the uncertain pixel with an uncertainty mask $m_t = \{m_{t,n}\}_{n=1}^{256 \times 256}$, where n indexes the pixel in the images, and $m_{t,n} \in \{0, 1\}, \forall t, n$:

$$\min_{\mathbf{w}, \mathbf{m}_t} \underbrace{\sum_{s \in S} \sum_{n=1}^{N} \| y_{s,n} - \tilde{y}_{s,n} \|_2^2}_{\mathscr{L}_{\text{reg}}^s (\mathbf{w})} + \underbrace{\sum_{t \in T} \sum_{n=1}^{N} \| (\hat{y}_{t,n} - \tilde{y}_{t,n}) m_{t,n} \|_2^2}_{\mathscr{L}_{\text{reg}}^t (\mathbf{w}, \mathbf{m}_t)} \tag{1}$$

$$s.t. \quad m_{t,n} = \begin{cases} 1 \; u_{t,n} < \epsilon \\ 0 \; u_{t,n} > \epsilon \end{cases}; \epsilon = min\{ \text{ top } p\% \text{ sorted } u_{t,n} \}, \tag{2}$$

where $x_{s,n}, y_{s,n}, x_{t,n}, \hat{y}_{t,n} \in [0, 255]$. For example, $y_{s,n}$ indicates the n-th pixel of the s-th source domain ground-truth cine MR image $\mathbf{y}_s$. $\tilde{y}_{s,n}$ and $\tilde{y}_{t,n}$ represent the generated source and target images, respectively. $\mathscr{L}_{reg}^s(\mathbf{w})$ and $\mathscr{L}_{reg}^t(\mathbf{w}, \mathbf{m}_t)$ are the regression loss of the source and target domain samples, respectively. Notably, there is only one network parameterized with $\mathbf{w}$, which is updated with the loss in both domains. $u_{t,n}$ is the to-be estimated uncertainty of a pixel and determines the value of the uncertainty mask $m_{t,n}$ with a threshold ϵ. ϵ is a critical parameter to control pseudo-label learning and selection, which is determined by a single meta portion parameter p, indicating the portion of pixels to be selected in the target domain. Empirically, we define ϵ in each iteration, by sorting $u_{t,n}$ in increasing order and set ϵ to minimum $u_{t,n}$ of the top $p \in [0, 1]$ percentile rank.

2.2 Bayesian Uncertainty Mask for Target Samples

Determining the mask value $m_{t,n}$ for the target sample requires the uncertainty estimation of $u_{t,n}$ in our self-training UDA. Notably, the lack of sufficient target domain data can result in the *epistemic* uncertainty w.r.t. the model parameters, while the noisy pseudo-label can lead to the *aleatoric* uncertainty [3, 11, 8].

To counter this, we model the *epistemic* uncertainty via Bayesian neural networks which learn a posterior distribution $p(\mathbf{w}/\mathbf{X}_T, \hat{\mathbf{Y}}_T)$ over the probabilistic model parameters rather than a set of deterministic parameters [25]. In particular, a tractable solution is to replace the true posterior distribution with a variational approximation $q(\mathbf{w})$, and dropout variational inference can be a practical technique. This can be seen as using the Bernoulli distribution as the approximation distribution $q(\mathbf{w})$ [5]. The K times prediction with independent dropout sampling is referred to as Monte Carlo (MC) dropout. We use the mean squared error (MSE) to measure the epistemic uncertainty as in [25], which assesses a one-dimensional regression model similar to [4]. Therefore, the epistemic uncertainty with MSE of each pixel with K times dropout generation is given by

$$u_{t,n}^{epistemic} = \frac{1}{K} \sum_{k=1}^{K} \| \tilde{y}_{t,n} - \mu_{t,n} \|_2^2; \; \mu_{t,n} = \frac{1}{K} \sum_{k=1}^{K} \tilde{y}_{t,n}, \tag{3}$$

where $\mu_{t,n}$ is the predictive mean of $\tilde{y}_{t,n}$.

Because of the different hardness and divergence and because the pseudo-label noise can vary for different $\mathbf{x}_t$, the heteroscedastic *aleatoric* uncertainty modeling is required [24, 13]. In this work, we use our network to transform $\mathbf{x}_t$, with its head split to predict both $\tilde{\mathbf{y}}_t$

and the variance map $\sigma_t^2 \in \mathbb{R}^{256 \times 256}$; and its element $\sigma_{t,n}^2$ is the predicted variance for the n-th pixel. We do not need "uncertainty labels" to learn σ_t^2 prediction. Rather, we can learn σ_t^2 implicitly from a regression loss function [13, 11]. The masked regression loss can be formulated as

$$\mathscr{L}_{reg}^t(\mathbf{w}, \mathbf{m}_t, \sigma_t^2) = \sum_{t \in T} \sum_{n=1}^{N} \left(\frac{1}{\sigma_{t,n}^2} \|(\hat{y}_{t,n} - \tilde{y}_{t,n})m_{t,n}\|_2^2 + \beta log \sigma_{t,n}^2 \right), \qquad (4)$$

which consists of a variance normalized residual regression term and an uncertainty regularization term. The second regularization term keeps the network from predicting an infinite uncertainty, i.e., zero loss, for all the data points. Then, the averaged aleatoric uncertainty of K times MC dropout can be measured by $u_{t,n}^{aleatoric} = \frac{1}{K}\sum_{k=1}^{K} \sigma_{t,n}^2$ [13, 11].

Moreover, minimizing Eq. (4) can be regarded as the Lagrangian with a multiplier β of. $\min_{\mathbf{w}} \sum_{t-T} \sum_{n=1}^{N} \frac{1}{\sigma_{t,n}^2} \|(\hat{y}_{t,n} - \tilde{y}_{t,n})m_{t,n}\|_2^2$; $s.t.$ $\sum_{t \in T} \sum_{n=1}^{N} log\sigma_{t,n}^2 < C^5$, where $C \in \mathbb{R}^{+}$ indicates the strength of the applied constraint. The condition term essentially controls the target domain predictive uncertainty, which is helpful for UDA [7]. Our final pixel-wise self-training UDA uncertainty $u_{t,n} = u_{t,n}^{\text{epistemic}} + u_{t,n}^{\text{aleatoric}}$ is a combination of the two uncertainties [11].

2.3 Training Protocol

As pointed out in [6], directly optimizing the self-training objectives can be difficult and thus the deterministic annealing expectation maximization (EM) algorithms are often used instead. Specifically, the generative self-training can be solved by alternating optimization based on the following **a)** and **b)** steps.

a) Pseudo-label and uncertainty mask generation.—With the current $\mathbf{w}$, apply the MC dropout for K times image translation of each target domain tagged MR image $\mathbf{x}_t$. We estimate the pixel-wise uncertainty $u_{t,n}$, and calculate the uncertainty mask $\mathbf{m}_t$ with the threshold ϵ. We set the pseudo-label of the selected pixel in this round as $\hat{y}_{t,n} = \mu_{t,n}$, i.e., the average value of K outputs.

b) Network w retraining.—Fix $\widehat{Y}_T = \{\hat{y}_t\}_{t=1}^T$, $\mathbf{M}_T = \{\mathbf{m}_t\}_{t=1}^T$ and solve:

$$\min_{\mathbf{w}} \sum_{s \in S} \sum_{n=1}^{N} \|y_{s,n} - \tilde{y}_{s,n}\|_2^2 + \sum_{t \in T} \sum_{n=1}^{N} \left(\frac{1}{\sigma_{t,n}^2} \|(\hat{y}_{t,n} - \tilde{y}_{t,n})m_{t,n}\|_2^2 + \beta log \sigma_{t,n}^2 \right) \qquad (5)$$

to update $\mathbf{w}$. Carrying out step **a)** and **b)** for one time is defined as one round in self-training. Intuitively, step **a)** is equivalent to simultaneously conducting pseudo-label learning and selection. In order to solve step **b)**, we can use a typical gradient method, e.g. Stochastic Gradient Descent (SGD). The meta parameter p is linearly increasing from 30% to 80% alongside the training to incorporate more pseudo-labels in the subsequent rounds as in [33].

3 Experiments and Results

We evaluated our framework on both cross-scanner and cross-center tagged-to-cine MR image synthesis tasks. For the labeled source domain, a total of 1,768 paired tagged and cine MR images from 10 healthy subjects at clinical center A were acquired. We followed the test time UDA setting [10], which uses only one unlabeled target subject in UDA training and testing.

For fair comparison, we adopted Pix2Pix [9] for our source domain training as in [20], and used the trained U-Net as the source model for all of the comparison methods. In order to align the absolute value of each loss, we empirically set weight $\beta = 1$ and $K = 20$. Our framework was implemented using the PyTorch deep learning toolbox. The GST training was performed on a V100 GPU, which took about 30 min. We note that K times MC dropout can be processed parallel. In each iteration, we sampled the same number of source and target domain samples.

3.1 Cross-scanner tagged-to-cine MR image synthesis

In the cross-scanner image synthesis setting, a total of 1,014 paired tagged and cine MR images from 5 healthy subjects in the target domain were acquired at clinical center A with a different scanner. As a result, there was an appearance discrepancy between the source and target domains.

The synthesis results using source domain Pix2Pix [9] without UDA training, gradually adversarial UDA (GAUDA) [2], and our proposed framework are shown in Fig. 2. Note that GAUDA with source domain initialization took about 2 hours for the training, which was four times slower than our GST framework. In addition, it was challenging to stabilize the adversarial training [1], thus yielding checkerboard artifacts. Furthermore, the hallucinated content with the domain-wise distribution alignment loss produced a relatively significant difference in shape and texture within the tongue between the real cine MR images. By contrast, our framework achieved the adaptation with relatively limited target data in the test time UDA setting [10], with faster convergence time. In addition, our framework did not rely on adversarial training, generating visually pleasing results with better structural consistency as shown in Fig. 2, which is crucial for subsequent analyses such as segmentation.

For an ablation study, in Fig. 2, we show the performance of GST without the aleatoric or epistemic uncertainty for the uncertainty mask, i.e., GST-A or GST-E. Without measuring the aleatoric uncertainty caused by the inaccurate label, GST-A exhibited a small distortion of the shape and boundary. Without measuring the epistemic uncertainty, GST-E yielded noisier results than GST.

The synthesized images were expected to have realistic-looking textures, and to be structurally cohesive with their corresponding ground truth images. For quantitative evaluation, we adopted widely used evaluation metrics: mean L1 error, structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), and unsupervised inception score (IS) [20]. Table 1 lists numerical comparisons using 5 testing subjects. The proposed GST

outperformed GAUDA [2] and ADDA [28] w.r.t. L1 error, SSIM, PSNR, and IS by a large margin.

3.2 Cross-center tagged-to-cine MR image synthesis

To further demonstrate the generality of our framework for the cross-center tagged-to-cine MR image synthesis task, we collected 120 tagged MR slices of a subject at clinical center B with a different scanner. As a result, the data at clinical center B had different soft tissue contrast and tag spacing, compared with clinical center A, and the head position was also different.

The qualitative results in Fig. 3 show that the anatomical structure of the tongue is better maintained using our framework with both the aleatoric and epistemic uncertainties. Due to the large domain gap present in the datasets between the two centers, the overall synthesis quality was not as good as the cross-scanner image synthesis task, as visually assessed. In Table 1, we provide the quantitative comparison using IS, which does not need the paired ground truth cine MR images [20]. Consistently with the cross-scanner setting, our GST outperformed adversarial training methods, including GAUDA and ADDA [2, 28], indicating the self-training can be a powerful technique for the generative UDA task, similar to the conventional discriminative self-training [33, 16].

4 Discussion and Conclusion

In this work, we presented a novel generative self-training framework for UDA and applied the framework to cross-scanner and cross-center tagged-to-MR image synthesis tasks. With a practical yet principled Bayesian uncertainty mask, our framework was able to control the confident pseudo-label selection. In addition, we systematically investigated both the aleatoric and epistemic uncertainties in generative self-training UDA. Our experimental results demonstrated that our framework yielded the superior performance, compared with the popular adversarial training UDA methods, as quantitatively and qualitatively assessed. The synthesized cine MRI with test time UDA can potentially be used to segment the tongue and to observe surface motion, without the additional acquisition cost and time.

Acknowledgements

References

1. Che T, Liu X, Li S, Ge Y, Zhang R, Xiong C, Bengio Y: Deep verifier networks: Verification of deep discriminative models with deep generative models. AAAI (2021)

2. Cui S, Wang S, Zhuo J, Su C, Huang Q, Tian Q: Gradually vanishing bridge for adversarial domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12455–12464 (2020)

3. Der Kiureghian A, Ditlevsen O: Aleatory or epistemic? does it matter? Structural safety 31(2), 105–112 (2009)

4. Fruehwirt W, Cobb AD, Mairhofer M, Weydemann L, Garn H, Schmidt R, Benke T, Dal-Bianco P, Ransmayr G, Waser M, et al. : Bayesian deep neural networks for low-cost neurophysiological markers of alzheimer's disease severity. arXiv preprint arXiv:1812.04994 (2018)

5. Gal Y, Ghahramani Z: Bayesian convolutional neural networks with bernoulli approximate variational inference. arXiv preprint arXiv:1506.02158 (2015)

6. Grandvalet Y, Bengio Y: Entropy regularization. (2006)

7. Han L, Zou Y, Gao R, Wang L, Metaxas D: Unsupervised domain adaptation via calibrating uncertainties. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 99–102 (2019)

8. Hu S, Worrall D, Knegt S, Veeling B, Huisman H, Welling M: Supervised uncertainty quantification for segmentation with multiple annotations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 137–145. Springer (2019)

9. Isola P, Zhu JY, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: CVPR. pp. 1125–1134 (2017)

10. Karani N, Erdil E, Chaitanya K, Konukoglu E: Test-time adaptable neural networks for robust medical image segmentation. Medical Image Analysis 68, 101907 (2021) [PubMed: 33341496]

11. Kendall A, Gal Y: What uncertainties do we need in bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)

12. Kumar MP, Packer B, Koller D: Self-paced learning for latent variable models. In: Advances in Neural Information Processing Systems. pp. 1189–1197 (2010)

13. Le QV, Smola AJ, Canu S: Heteroscedastic gaussian process regression. In: Proceedings of the 22nd international conference on Machine learning. pp. 489–496 (2005)

14. Liu X, Fan F, Kong L, Diao Z, Xie W, Lu J, You J: Unimodal regularized neuron stick-breaking for ordinal classification. Neurocomputing 388, 34–44 (2020)

15. Liu X, Hu B, Jin L, Han X, Xing F, Ouyang J, Lu J, El Fakhri G, Woo J: Domain generalization under conditional and label shifts via variational bayesian inference. In: IJCAI (2021)

16. Liu X, Hu B, Liu X, Lu J, You J, Kong L: Energy-constrained self-training for unsupervised domain adaptation. ICPR (2020)

17. Liu X, Liu X, Hu B, Ji W, Xing F, Lu J, You J, Kuo CCJ, Fakhri GE, Woo J: Subtype-aware unsupervised domain adaptation for medical diagnosis. AAAI (2021)

18. Liu X, Xing F, El Fakhri G, Woo J: Adapting off-the-shelf source segmenter for target medical image segmentation. In: MICCAI (2021)

19. Liu X, Xing F, El Fakhri G, Woo J: A unified conditional disentanglement framework for multimodal brain mr image translation. In: ISBI. pp. 10–14. IEEE (2021)

20. Liu X, Xing F, Prince JL, Carass A, Stone M, Fakhri GE, Woo J: Dual-cycle constrained bijective VAE-GAN for tagged-to-cine magnetic resonance image synthesis. ISBI (2021)

21. Liu X, Xing F, Yang C, Kuo CCJ, El Fakhri G, Woo J: Symmetric-constrained irregular structure inpainting for brain mri registration with tumor pathology. In: Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. BrainLes (Workshop). vol. 12658, p. 80 (2021)

22. Liu X, Zou Y, Song Y, Yang C, You J, K Vijaya Kumar B: Ordinal regression with neuron stick-breaking for medical diagnosis. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018)

23. Mei K, Zhu C, Zou J, Zhang S: Instance adaptive self-training for unsupervised domain adaptation. ECCV (2020)

24. Nix DA, Weigend AS: Estimating the mean and variance of the target probability distribution. In: Proceedings of 1994 ieee international conference on neural networks (ICNN'94). vol. 1, pp. 55–60. IEEE (1994)

25. Rasmussen CE: Gaussian processes in machine learning. In: Summer school on machine learning. pp. 63–71. Springer (2003)

26. Shin I, Woo S, Pan F, Kweon IS: Two-phase pseudo label densification for self-training based domain adaptation. In: European Conference on Computer Vision. pp. 532–548. Springer (2020)

27. Tang K, Ramanathan V, Fei-Fei L, Koller D: Shifting weights: Adapting object detectors from image to video. In: NIPS (2012)

28. Tzeng E, Hoffman J, Saenko K, Darrell T: Adversarial discriminative domain adaptation. In: CVPR (2017)

29. Wang J, Liu X, Wang F, Zheng L, Gao F, Zhang H, Zhang X, Xie W, Wang B: Automated interpretation of congenital heart disease from multi-view echocardiograms. Medical Image Analysis 69, 101942 (2021) [PubMed: 33418465]

30. Wang M, Deng W: Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153 (2018)

31. Wei C, Shen K, Chen Y, Ma T: Theoretical analysis of self-training with deep networks on unlabeled data. arXiv preprint arXiv:2010.03622 (2021)

32. Zhu X: Semi-supervised learning tutorial. In: ICML tutorial (2007)

33. Zou Y, Yu Z, Liu X, Kumar B, Wang J: Confidence regularized self-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5982–5991 (2019)
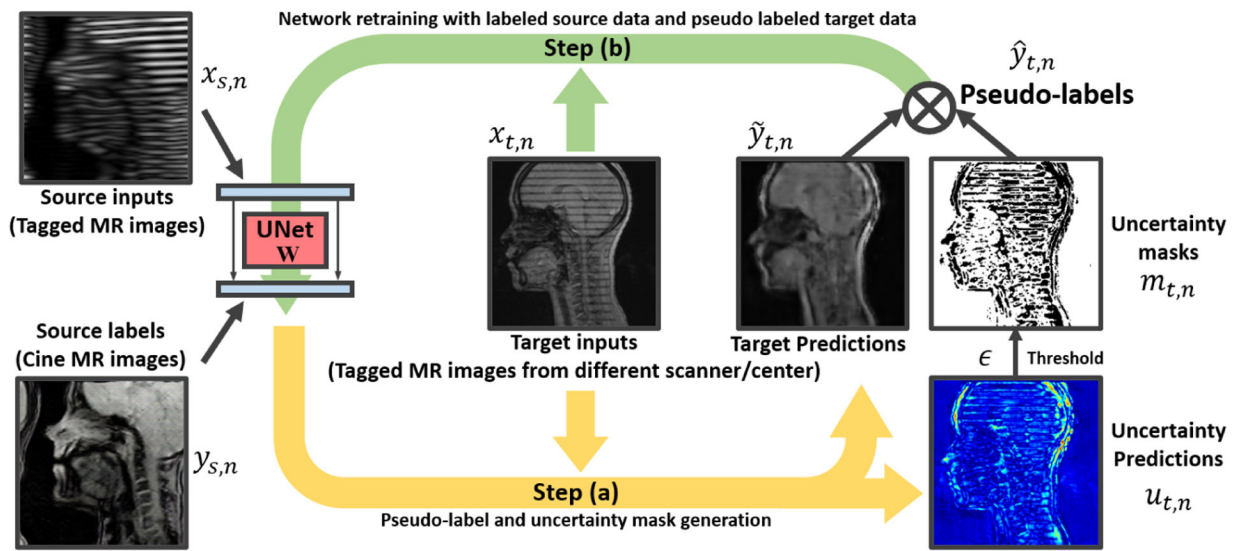
Fig. 1:

Illustration of our generative self-training UDA for tagged-to-cine MR image synthesis. In each iteration, two-step alternative training is carried out.
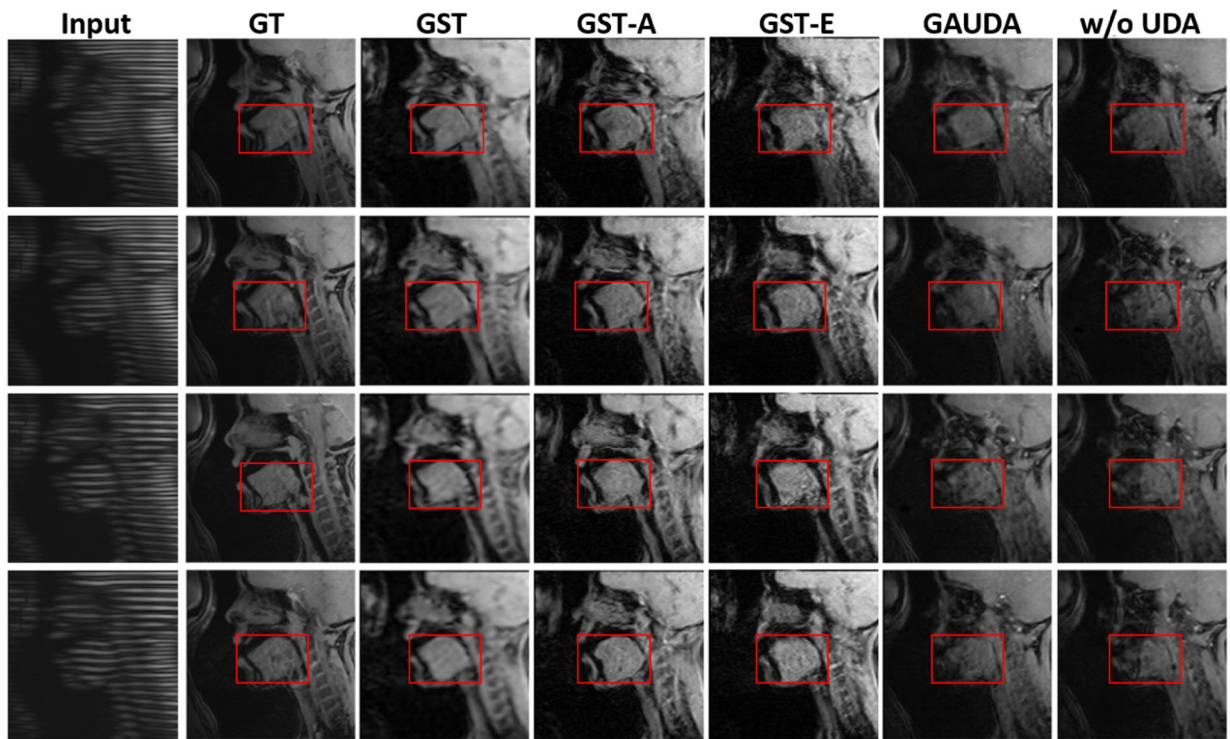
Fig. 2:
Comparison of different UDA methods on the cross-scanner tagged-to-cine MR image synthesis task, including our proposed GST, GST-A, and GST-E, adversarial UDA [2]*, and Pix2Pix [9] without adaptation. * indicates the first attempt at tagged-to-cine MR image synthesis. GT indicates the ground-truth.
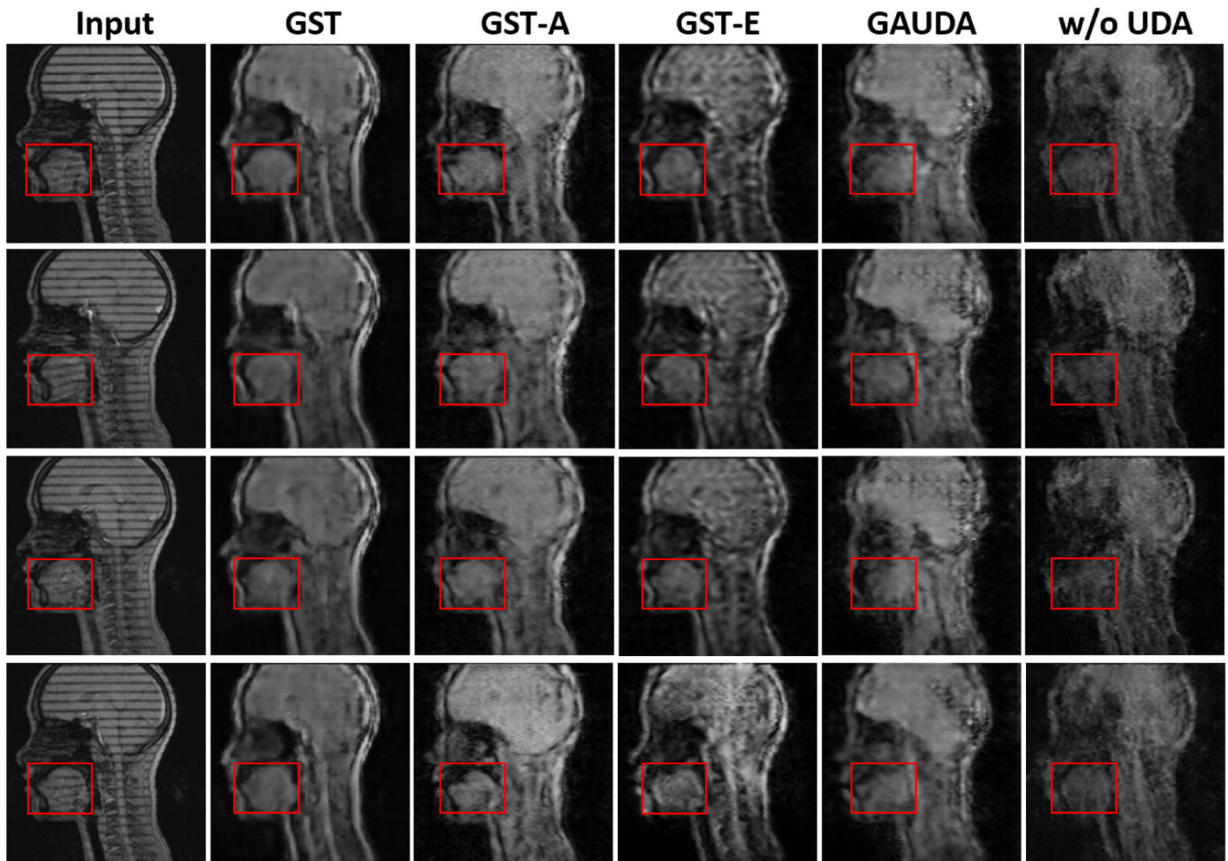
Fig. 3:

Comparison of different UDA methods on the cross-center tagged-to-cine MR image synthesis task, including our proposed GST, GST-A, and GST-E, adversarial UDA [2]*, and Pix2Pix [9] without adaptation. * indicates the first attempt at tagged-to-cine MR image synthesis.

Table 1:

Numerical comparisons of cross-scanner and cross-center evaluations.

Methods	Cross-scanner				Cross-center
	L1 ↓	SSIM ↑	PSNR ↑	IS ↑	IS ↑
w/o UDA [9]	176.4±0.1	0.8325±0.0012	26.31±0.05	8.73±0.12	5.32±0.11
ADDA [28]	168.2±0.2	0.8784±0.0013	33.15±0.04	10.38±0.11	8.69±0.10
GAUDA [2]	161.7±0.1	0.8813±0.0012	33.27±0.06	10,62±0.13	8.83±0.14
GST	**158.6**±0.2	**0.9078**±0.0011	**34.48**±0.05	**12.63**±0.12	**9.76**±0.11
GST-A	159.5±0.3	0.8997±0.0011	34.03±0.04	12.03±0.12	9.54±0.13
GST-E	159.8±0.1	0.9026±0.0013	34.05±0.05	11.95±0.11	9.58±0.12

± standard deviation is reported over three evaluations.